

Energy Efficiency Benefits of RAN-as-a-Service Concept for a Cloud-Based 5G Mobile Network Infrastructure

DARIO SABELLA¹, ANTONIO DE DOMENICO², EFSTATHIOS KATRANARAS³, (Member, IEEE), MUHAMMAD ALI IMRAN³, (Senior Member, IEEE), MARCO DI GIROLAMO⁴, UMER SALIM⁵, MASSINISSA LALAM⁶, KONSTANTINOS SAMDANIS⁷, AND ANDREAS MAEDER⁷

¹Telecom Italia, Turin 10148, Italy

²Laboratory for Electronics and Information Technology, Atomic Energy Commission, Grenoble 38054, France

³University of Surrey, Guildford GU2 7XH, U.K.

⁴HP Italy Innovation Center, Milan 20003, Italy

⁵Intel Mobile Communications, Paris 75017, France

⁶Sagemcom Broadband, Paris 92848, France

⁷NEC Laboratories Europe, Heidelberg 69115, Germany

Corresponding author: D. Sabella (dario.sabella@telecomitalia.it)

This work was supported by the European Community's 7th Framework Programme (FP7/2007-2013) under Grant 317941 through the iJOIN Project.

ABSTRACT This paper focuses on energy efficiency aspects and related benefits of radio-access-network-as-a-service (RANaaS) implementation (using commodity hardware) as architectural evolution of LTE-advanced networks toward 5G infrastructure. RANaaS is a novel concept introduced recently, which enables the partial centralization of RAN functionalities depending on the actual needs as well as on network characteristics. In the view of future definition of 5G systems, this cloud-based design is an important solution in terms of efficient usage of network resources. The aim of this paper is to give a vision of the advantages of the RANaaS, to present its benefits in terms of energy efficiency and to propose a consistent system-level power model as a reference for assessing innovative functionalities toward 5G systems. The incremental benefits through the years are also discussed in perspective, by considering technological evolution of IT platforms and the increasing matching between their capabilities and the need for progressive virtualization of RAN functionalities. The description is complemented by an exemplary evaluation in terms of energy efficiency, analyzing the achievable gains associated with the RANaaS paradigm.

INDEX TERMS Energy efficiency, wireless communication, radio access networks, RAN-as-a-service, power model, cloud-RAN, LTE-advanced, 5G.

I. INTRODUCTION

The current vision towards 5G is often driven by traffic forecasts that suggest increasing data volumes, number of more intelligent terminals and an ever growing capacity and service-aware demand. This discussion is also accompanied by the fundamental question related to the need to define a new air interface or consider 5G as an evolution of current systems.

Launching a higher number of base stations, i.e. introducing traditional network densification, may provide a solution for such a tremendous traffic increase, but would result in increased network infrastructure costs, which are expected to see a consequent explosion through the years. In particular, energy consumption is an important part

of Operational Expenditure (OPEX) and its relevance in upcoming systems will be somehow proportional to network growth. In fact, energy consumption at network level depends on large part on the number of installed radio base stations. In addition to this essential densification of network nodes, integrating new systems upon the existing ones unavoidably increases the energy consumption, even if new systems are more efficient than the old ones (this happened by adding LTE on top of 2G/3G). Increased energy consumption means higher costs and a greater carbon footprint, since today mobile systems are present everywhere in the world. The European Commission (EC) recognized the need for further actions towards energy efficiency and green communications and introduced the Code-of-Conduct [1] to provide a policy

that regulates energy consumption and carbon dioxide emissions.

As a consequence, Energy Efficiency (EE) and sustainability of 5G networks have recently received significant attention from mobile operators, vendors and research projects [2]. Figure 1 shows our vision of the EE evolution in mobile networks toward a sustainable 5G, where the exponential mobile traffic growth toward 2020 (blue curve) goes with a stable network energy consumption (red curve), resulting in an increasing EE of the system through the years (green curve).

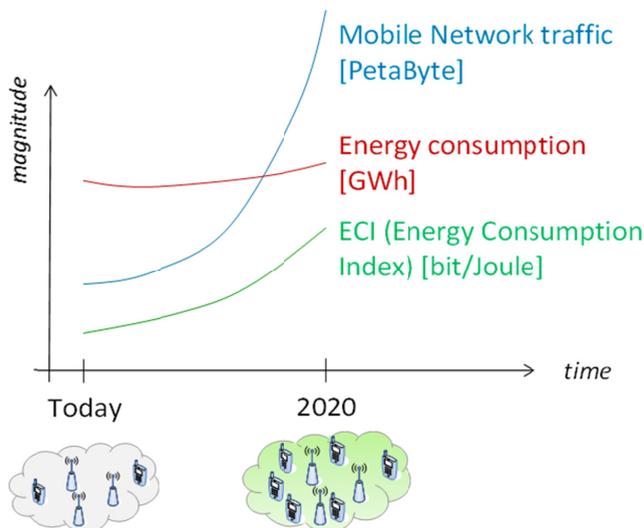


FIGURE 1. Energy efficiency evolution in mobile networks toward a sustainable 5G.

From the standardization point of view, EE in wireless systems mainly concentrates on the efforts of 3GPP for LTE and LTE-Advanced (LTE-A) [3]. The 3GPP Radio Access Network (RAN) groups were the first to deal with EE in Release 9 by investigating energy-aware network management with small cells, allowing LTE base stations (called eNBs) to switch-off, based on local load information and configuration data or via the Operation and Management (OAM). The RAN groups further introduced the “cell DTX” mechanism, which deactivates the transmitter of an eNB based on certain patterns, conserving energy on the fraction of inactivity time [4]. Energy saving management was later suggested by the Service Architecture Group 5 (SA5) that introduced Self-Organizing Network (SON) operations in relation with overlaid and dense urban networks, wherein certain eNBs may compensate in terms of coverage and service allowing selected eNBs to be powered-off [5].

Besides 3GPP, the effort of the research community in this topic is also notable. Current activities mainly focus on evolving flexible hardware for enhancing eNBs, novel architectures based on small cells deployment and adaptive schemes that adjust network capacity with respect to service loads [2]. Nevertheless, while research on green communications has produced notable results, improving the wireless

network EE is still an open research field with GreenTouch recently announcing that the potential of reducing the net energy consumption may reach up to 90% by 2020 [6]. However, a main challenge is to elaborate realistic and complete models of innovative solutions for new generation networks, in order to correctly influence the definition of affordable and sustainable 5G systems.

In [7], the EE benefits of a centralized RAN are analyzed in terms of hardware design, i.e., cooling and soft technologies including cooperative processing, virtualization and dynamic cell re-configuration. In addition, signaling and control optimizations are explored to move away from the conventional connection-oriented paradigm, i.e., the use of “virtual” eNBs with irregular antenna arrays and new interference mitigation schemes.

A more flexible cloud-based RAN architecture is proposed in [8], where the front-haul is logically re-adapted to the corresponding requirements. Such cloud-RAN architecture enables energy saving on the cloud-RAN platform, by re-arranging the number of active Base Band processing Units (BBUs), when the traffic is low.

More recently, the concept of RAN-as-a-Service (RANaaS) has been introduced in [9], where some RAN functionalities are partially and flexibly centralized, depending on the actual load and network characteristics. In the view of future definition of 5G systems, these cloud-based designs together with RAN sharing mechanisms appear to be the most promising solutions in terms of efficient usage of network resources.

The aim of the paper is to present a vision of the advantages of the RANaaS paradigm and its benefits in terms of EE. To do so, we will introduce a consistent system-level power consumption model, which is intended to be a useful reference for the EE performance evaluation of innovative RAN mechanisms. The rest of the paper is organized as follows. Section II describes the proposed mobile access architecture. Section III gives an overview of current status and future trends of IT platforms, with a description of related performances offered for hosting RANaaS entities. Section IV provides a detailed analysis of the proposed architecture from an EE perspective, with a derivation of a comprehensive power model for each involved entity. Section V gives an exemplary description of small cell management using the RANaaS paradigm, as a promising solution for future sustainable 5G systems. Finally, Section VI concludes the paper.

II. REFERENCE SCENARIO AND ARCHITECTURE

In mobile communications, the trend has always been to push the computation burden toward the last miles, to reduce the round-trip-time and improve the system reactivity. With dense small cells, coordination is needed again to deal with the high level of interference introduced due to the proximity of base stations. To cope with this requirement, we consider in Figure 2 an evolutionary architecture of the LTE-A one [10], where RAN functions associated with small cells (iSCs) can be centralized in a flexible manner. Such function centralization reduces the processing on radio access, where

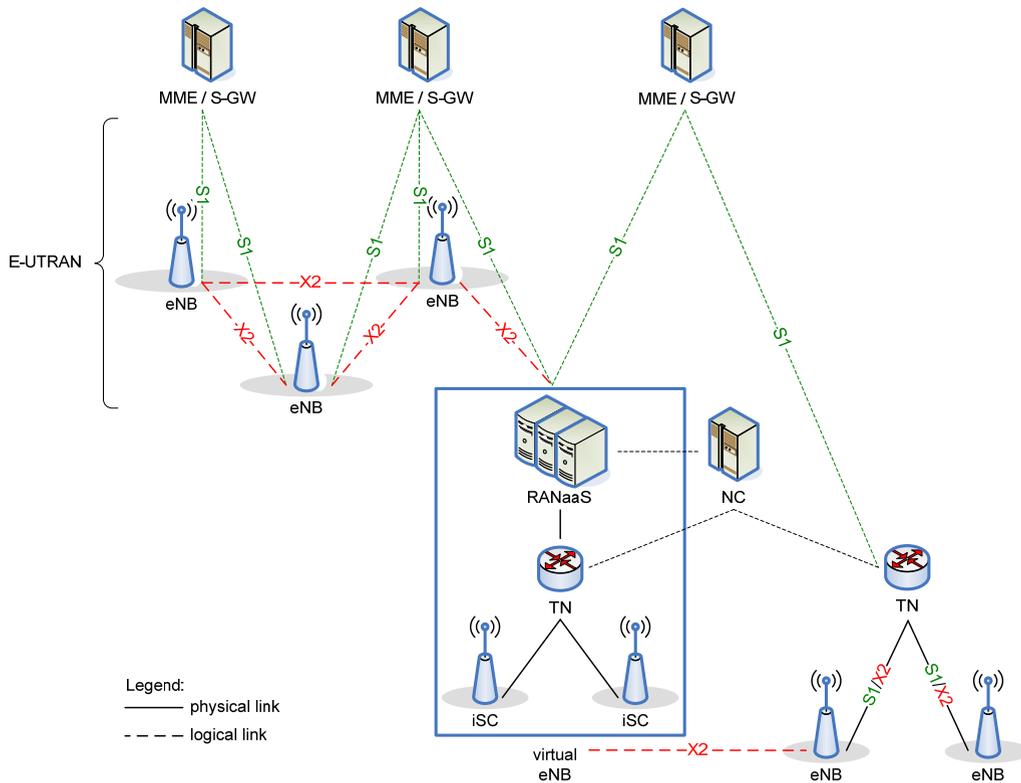


FIGURE 2. Mobile network architecture implementing the RANaaS concept [9].

iSCs with low transmission power are used to allow high data-rates and enhanced EE compared to macro cells. Moreover, to jointly optimize RAN and backhaul, a Network Controller (NC), i.e., a software defined network-based controller, configures the routing among the Transport Nodes (TNs) based on the associated constraints such as RAN/backhaul load, user density and mobility.

The purpose of the RANaaS, based on generic data centres, is to centrally execute part of the RAN functionalities, thus benefitting from centralization gains, which are fundamental in ultra-dense deployments. In practice, the RANaaS and iSCs entities appear as classical eNBs to the existing network. Therefore, such “virtual eNB” (veNB) entity can be seamlessly integrated in the legacy architecture.

Under such setup, the power consumption of the veNB will take into account the iSCs, the backhaul network, and the RANaaS. At a first glance, it appears that this power consumption will be much higher than those required by a classical deployment that covers the same area, but this will have to be evaluated relatively to the effective gain on the system performance as well. For instance, this architecture is potentially enabling advanced RAN sharing solutions that may dramatically improve the system performance. Accordingly, EE evaluations are traditionally performed by considering two kinds of metrics (and variations of them) [2]: Energy per Information Bit [J/bit] and Power per Area Unit [W/m²].

III. EVOLUTION OF CLOUD INFRASTRUCTURES AND RELATED PERFORMANCE

The cloud computing model is the computational paradigm which is substantially marking the latest and future era in the IT realm. According to one of the most recognized definitions, the one elaborated by NIST [11], cloud computing is:

“...a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction...”

Although standard cloud computing is not the only possible way to implement the RANaaS model described in chapter II, it is the most obvious “first option” to take into account, since its characteristics (like full transparency of underlying physical resources, potential portability and dynamic scalability) are optimally fitting the main drivers behind RAN virtualization and centralization. On the other hand, sticking to an industry standard model like cloud computing imposes some “overheads” requested to preserve generality, which can have an impact on the overall performance and ultimately on the feasibility of RAN centralization. Assessing such constraints and limitations is among the key objectives of research initiatives underlying what is discussed in the current paper.

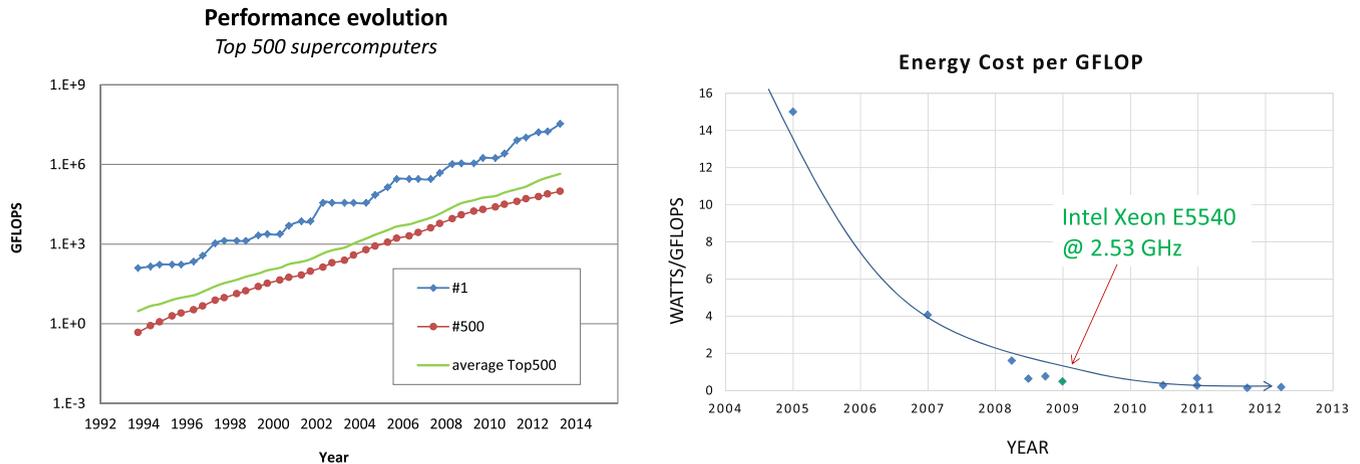


FIGURE 3. Evolution of computational power (GFLOPS) (left) and Energy Efficiency (Watts/GFLOPS) (right).

The RANaaS concept sets challenging requirements to the hardware (and software) infrastructure on which the RAN functions have to be outplaced from standard LTE (LTE-A) base stations. On one hand, the infrastructure should be based on industry standard servers (ISS) and computational equipment (such as storage and network appliances), to keep a reasonable cost balance with respect to standard base stations: paying a financial penalty would override one of the key reasons to go after centralization, i.e. CAPEX reduction. On the other hand, commoditization must not hinder the minimal performance requested to sustain the performance requirements of the processing and management functions swapped to the RANaaS layer. Finally, centralization must be kept under control from an EE standpoint, to make sure that the centralization gains of functional split do not introduce a total excessive energy consumption increase at overall system level.

Accordingly, one of the key objectives of our research is finding out a set of minimal conditions that guarantee performance and efficiency goals at once in acceptable extent. One of such assessments concerns the RANaaS hardware and software infrastructure: ISS have gone through a relevant evolution over the last years, affecting their performance, the performance/footprint ratio and the EE. Figure 3 gives a statistical visual idea of such trend, considering the top 500 supercomputers. Supercomputers are not the actual target of RAN centralization, for the reasons exposed above orienting the research towards ISS equipment. However, ISS servers are not benchmarked versus low-level operations/second Key Performance Indicators (KPIs); their standard industry benchmarks are executed through specific applications, not easily relatable to RAN functions. Nevertheless, the figure trend can very well approximate a similar evolution undergone in the ISS domain.

In detail, the left-side of Figure 3 fully captures the growth in computational power. On the other hand,

right-side of Figure 3 depicts energy cost per billion of FLOPS (Floating point Operations Per Second) and shows how deployment of huge servers, realizing trillions of operations, is more and more becoming financially and environmentally sustainable. These aspects are also at some extent interrelated, e.g., a reduced footprint can also help to save power.

The current standard generation mostly employed in large data centres is based on X86 servers, in rackable or blade form factor. However, the ISS technology is undergoing a breakthrough transformation, which is bringing towards much more powerful servers with a far smaller form factor and a previously unknown reconfigurability level: *Hyperscale* servers have larger computational power, along with a high degree of scalability, vertical or horizontal (this latter typical enabler of the cloud computing model). Also, new processor families like ARM or Atom are starting to be employed inside ISS, hugely increasing the computational power density per square meter and, thus, reducing the facility costs of a data-centre.

Finally, these new systems implement the *software defined server* concept to optimize their configuration to the peculiarities of the workload to be executed on them. For instance, HP Moonshot chassis are 7.5" tall and can host 45 hot-plug independent server cartridges, sharing power, cooling and management resources [12]. Server cartridges are based on Intel Atom processors and are design-optimized for specific classes of applications. A 45-server chassis shows a power consumption of around 850W, able to power 180×2.0 GHz threads, with 2GB of RAM for each thread, at under 5W per thread. Cartridges can be optimized by adjusting local mass storage capacity on each module and by experimenting with chassis-level network switch configurations, for instance to focus on front-end performance. With similar server architectures, it is easy to envision an upcoming availability of standard servers optimized on a specific problem domain like RAN processing.

Clearly, this breakthrough passage in the ISS technology has relevant impacts on the applicability and actual viability of the RANaaS concept. The above aspects drive a bulky cost reduction of computational power in data centres, on one hand filling the performance gap existing between custom components and ISS servers, on the other cutting costs and making an ISS based infrastructure really competitive versus legacy eNBs.

IV. POWER CONSUMPTION MODEL FOR CLOUD-BASED MOBILE NETWORKS

While the traditional architecture considers several sophisticated small cells, our reference system is composed by several iSCs of low processing capability and a RANaaS platform where pooling of complex (e.g. baseband) processing can be performed. In order to investigate the convenience (from an energy performance perspective) of the proposed architecture compared to the traditional one, we need to introduce a generalized holistic power model; in fact, power consumption at system level should be evaluated by considering the sum of all power consumption contributions in the network. This will help us to perform a quantitative analysis on the power consumption of the RANaaS concept in cloud-based mobile networks and evaluate the potential benefits in terms of energy efficiency in different load conditions.

Considering the system architecture as introduced in Section 2, comprising N_{iSC} iSCs in total, its overall power consumption can be modelled by:

$$P_{Total} = P_{RANaaS} + P_{Bh} + \sum_{n=1}^{N_{iSC}} P_{iSC-n} \quad (1)$$

where P_{RANaaS} , P_{Bh} , and P_{iSC-n} stand for the power consumed at the RANaaS platform, the power needs for backhauling network, and the power usage at any iSCn n , respectively.

In the following, the power consumption of each individual network element is discussed. Furthermore, some examples of measures are provided to correlate and obtain an idea on the order of magnitude of each element's power consumption, depending on the cells' load (which is interrelated to the cells' RF output power).¹

A. iSC POWER CONSUMPTION

The FP7 EARTH has investigated how the power consumption of distinct components of several eNBs, such as power amplifier (PA), baseband engine (BB), Radio Frequency (RF) small-signal transceiver, direct-current (DC)-DC converter, main supply (MS), and active cooling (CO), depends on the transmission bandwidth, the transmission power, and the number of radio chains/antennas [13]. According to

¹In addition, this paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the authors. This includes: a multimedia video endorsing the importance of energy efficiency assessment applied to architectures based on the RANaaS concept, the Excel implementation of the overall power consumption model for RANaaS-enabled mobile networks, and a readme file. This material is about 2 MB in size.

this study, the maximum power consumption of a single eNB sector can be given by [14]:

$$P_{eNB} = \frac{P_{BB} + P_{RF} + P_{PA}}{(1 - \sigma_{DC})(1 - \sigma_{MS})(1 - \sigma_{CO})} \quad (2)$$

where P_{BB} , P_{RF} , P_{PA} and σ_{DC} , σ_{MS} , σ_{CO} denote the power consumption and loss factors, respectively, of the different components. Furthermore, it was found that an affine function of the transmission power (comprising a static and a linearly increasing load-dependent share) can approximate very well the generalized model of eNBs [15], i.e.,:

$$P_{eNB} = \begin{cases} P_0 + y \cdot \Delta_p \cdot P_{max} & \text{if } 0 < y < 1 \\ P_{sleep} & \text{if } y = 0 \end{cases} \quad (3)$$

where P_0 , P_{max} and P_{sleep} stand for the eNB power consumption at zero load, full load and sleep mode (considering that eNBs may enter a low consumption sleep mode where some of their main units are turned off when no data is received or transmitted), respectively. Furthermore, y and Δ_p denote the cell load and the slope of the load-dependent linear model.

To approximate iSC power consumption, we have adopted and combined the aforementioned models and we have taken into account that: 1) the BB and RF power consumptions scale linearly with the system bandwidth (W) and the number of antennas (N_{ant}) used [16], 2) PA power consumption can be approximated as a linear function of the PA output power, and 3) no active cooling is needed in that case. Moreover, for simplicity, we have considered that only the PA is turned off when an iSC goes to sleep mode (i.e. no BB engine reductions due to sleep mode are considered). Therefore, the power consumption of an iSC n with cell load y_n can be given by:

$$P_{iSC-n} = \frac{N_{ant} \frac{W}{10[\text{MHz}]} (P'_{BB} + P'_{RF}) + y_n P_{PA-max}}{(1 - \sigma_{DC})(1 - \sigma_{MC})} \quad (4)$$

where P'_{BB} and P'_{RF} denote the BB and RF base consumptions (i.e. using 10MHz and one antenna) while P_{PA-max} is the PA maximum transmission power.

In general, we consider that an iSC may implement only a portion of the eNB protocol stack. Therefore, its power consumption will be lower and upper bounded by the two extreme cases: a) Radio Remote Head (RRH) and b) baseline Small Cell, respectively. RRHs are considered as low complexity and processing nodes that solely perform RF operations and rely on self-backhauling (i.e. $P_{BB} = 0$ in that case). On the other hand, baseline small cells perform all the BB operations.

Table 1 provides exemplary realistic values for the parameters associated to the iSC power model, considering two (per-antenna) maximum transmit power, i.e., 24 dBm ($P_{Tx,1}$) and 30 dBm ($P_{Tx,2}$), and Figure 4 depicts the estimate power consumption of iSCs.

TABLE 1. iSC power consumption - exemplary realistic parameter values [13].

iSC			
System Bandwidth (W)	10 MHz	Main Supply losses (σ_{MC})	7.7 %
# antennas per iSC (N_{ant})	2	DC-DC conversion losses (σ_{DC})	6.4 %
BB base consumption (P'_{BB})	0 for RRH 6.8 W for baseline SC	PA max consumption (P_{PA-max})	0.8 W if $P_{Tx,1}$ 3.2 W if $P_{Tx,2}$
RF base consumption (P'_{RF})	0.8W if $P_{Tx,1}$ 1.5W if $P_{Tx,2}$	Load of cell n (y_n)	0 – 100 %

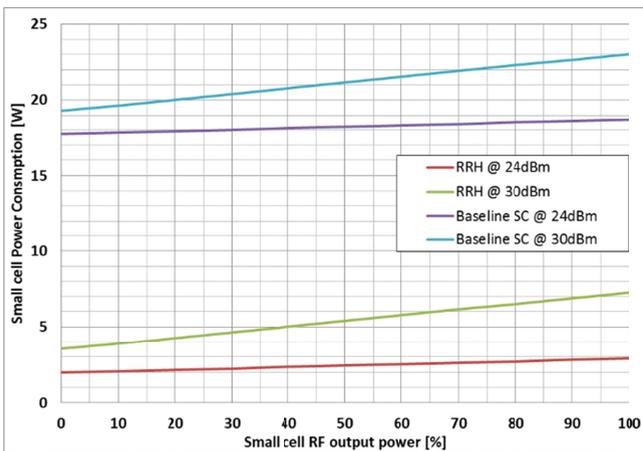


FIGURE 4. Complete small cell and RRH power consumption with respect to different RF output power and constraints.

B. RANaaS PLATFORM POWER CONSUMPTION

In order to obtain an accurate estimation on the power consumption of the RANaaS platform due to BB processing² moved from iSCs, we make use of the commodity hardware consumption coming from the IT world; Fit4Green has investigated the power consumption for IT resources of data centres [17]. In particular, results for the various computing style servers are provided using a monitoring tool and a generic power consumption prediction model. Considering the measurements results for the cloud computing testbed hardware equipment as the closest paradigm for the BBUs at a RANaaS platform, it can be observed that a linear model can approximate well a server’s power consumption (P_{srv}) versus its percentage CPU usage (x_{srv} %)

$$P_{srv} = P_0^{srv} + \Delta_p^{srv} P_{max}^{srv} x_{srv} \quad (5)$$

²Note that for fair comparison, we consider as power consumption of the RANaaS platform only the extra power needed for computations due to BB processing moved to RANaaS from iSCs (e.g. at a RANaaS BBU). Other functionalities at RANaaS platform can be assumed that exist also at the core network of a conventional system.

where P_0^{srv} and P_{max}^{srv} denote the power consumption of the server when in idle mode and maximum usage, respectively, while Δ_p^{srv} stands for the slope of the equivalent linear power model which depends on the specific server considered.

Considering the RANaaS BBU as an enclosure hosting several identical interconnected ISS Blade servers (each considered as a set of multiple processors) equally sharing the requested workload, the servers’ processing capacity (X_{Cap}), in Giga-Operations-per-Second³ (GOPS), will define the total number of servers (N_{srv}) required to process the system BB-related workload (X in GOPS) moved to RANaaS BBU, i.e.:

$$N_{server} = \left\lceil \frac{X}{X_{Cap}} \right\rceil \quad (6)$$

and the percentage CPU usage at each server will be:

$$x_{server} = \frac{X}{X_{Cap} N_{server}} \cdot 100\% \quad (7)$$

The question that arises next is how the extra RANaaS BBU workload can be related to cell load. In that regard, the work in [16] considered the functionalities of various base station types and examined how the GOPS per function block scale with cell load for a specific reference system. Targeting to a more generalized view, Werthman et al. defined recently the resource effort required to serve a UE at a specific time as a function of the number of used antennas, the modulation bits, the code rate, the number of spatial MIMO-layers and the number of allocated frequency/time resources each as allocated to the UE at that time [18]. Based on this work, we introduce an average sum to approximate the total extra RANaaS BBU workload required to serve all UEs when an average β_{BB} % of BB processing is assumed to be moved towards RANaaS platform from each iSC. To this end, the GOPS required at RANaaS platform will depend on the total number of iSCs in the system, the load of each iSC, the

³It is noted that the processing capacity of the server is usually expressed in GFLOPS [19]; however, it can be converted in GOPS, and in this work we use a 1:1 ratio as a conservative estimation.

TABLE 2. RANaaS platform power consumption - exemplary realistic parameter values.

RANaaS platform			
Server Capacity (X_{Cap})	324 GFLOPS [19]		
Server idle consumption (P_0^{srv})	120 W	Consumption at Server Max Workload ($P_{\text{max}}^{\text{srv}}$)	215 W
Linear model slope (Δ_p^{srv})	0.44	# iSCs in cluster (N_{iSC})	10
Average # of antennas used to serve a UE (N_{Tx})	2	% of BB processing moved into RANaaS from each iSC (β_{BB})	10 50 100 %
Average # of data bits per symbol per UE (e_{MCS})	4/3	Average # of spatial MIMO layers used per UE (e_{MIMO})	1.1

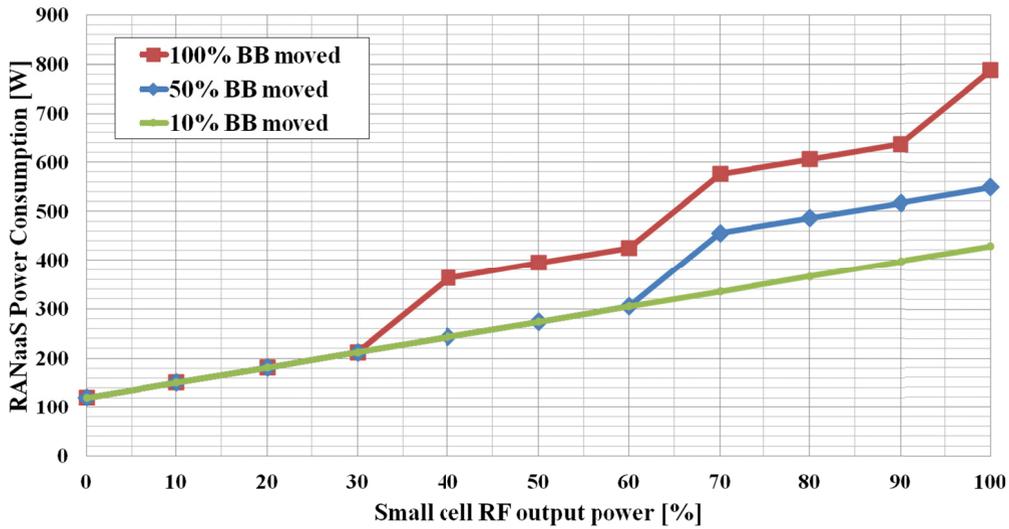


FIGURE 5. RANaaS power consumption with respect to the small cell RF output power for different BB shift options.

system bandwidth, the number of antennas used to serve a UE per iSC (N_{Tx}), the average number of data bits per symbol per user (e_{MCS}) and the average number of MIMO layers per user (e_{MIMO}):

$$X = \frac{\beta_{\text{BB}}}{100} \cdot \frac{W}{10 [\text{MHz}]} \left(30N_{\text{Tx}} + 10N_{\text{Tx}}^2 + 20 \frac{e_{\text{MCS}}}{6} e_{\text{MIMO}} \right) \times \sum_{n=1}^{N_{\text{iSC}}} y_n \quad (8)$$

Therefore, for any RANaaS BBU comprising servers with specific processing capacity, its overall power consumption

due to BB processing moved towards RANaaS platform will be the sum of the power consumption at each of the required servers (9), as shown at the bottom of this page.

To obtain an exemplary view on the power consumption of RANaaS platform we consider a cloud computing server comprising Intel Xeon E5540 processors as presented in [17] under the high efficiency configuration policy (i.e. CPU frequency always set to maximum value). Table 2 introduces the respective realistic values and Figure 5 illustrates the power consumption of this specific RANaaS BBU with respect to the small cell RF output power for different BB shift.

$$P_{\text{RANaaS}} = \left[\frac{X(y)}{X_{\text{Cap}}} \right] \cdot \left(P_0^{\text{srv}} + \sum_{n=1}^{N_{\text{iSC}}} y_n \cdot \frac{W}{10\text{MHz}} \cdot \frac{(30N_{\text{Tx}} + 10N_{\text{Tx}}^2 + 20 \frac{e_{\text{MCS}}}{6} e_{\text{MIMO}})}{X_{\text{Cap}} \left[\frac{X(y)}{X_{\text{Cap}}} \right]} \Delta_p^{\text{srv}} P_{\text{max}}^{\text{srv}} \right). \quad (9)$$

TABLE 3. Backhauling network power consumption - exemplary realistic parameter values.

Backhauling Network			
# microwave antennas per iSC (N_{mw})	2	Switch maximum capacity (C_{switch})	36 Gbps
Switch basic consumption (P_s)	53 W	Average cell capacity (Y_{max})	86.4 Mbps
Cell load to backhaul traffic factor ($f_{cell-Bh}$)	128 %	Traffic threshold between low/high power regions (C_{thr})	500 Mbps
Node power region for idle/low/high traffic conditions ($P_{idle/low/high-traffic}$)	22.2 / 37 / 92.5 W		

As we can see in Figure 5, RANaaS power consumption trend depends on the particular functional split considered (and the consequent percentage of BB processing moved in the RANaaS). Linear increase of the power consumption is obtained (with 10% BB in RANaaS) when the processing capacity of a single server is sufficient to handle all load in the small cell. When more computational effort is moved in the RANaaS (e.g. 50%, 100%), the power consumption trend shows some jumps due to the fact that more servers are required to handle the processing workload moved at the RANaaS as the cell load and % shift increases.

C. BACKHAUL POWER CONSUMPTION

The last important element that we have modelled is the backhaul network. In general, centralised systems have notable backhaul load; therefore, power consumption due to data transport and switching can become a significant percentage of the total system power consumption [2].

Monti et al. provided some fundamental power consumption models for data transport through various backhaul technologies and topologies in small cells [20]. Considering self-backhauling iSCs with microwave links and omitting TNs for simplicity, the backhaul power consumption can be estimated by adjusting the respective model in [20] to our system architecture paradigm. In our case, backhaul power consumption shall scale with the number of iSCs in the system, the load of each iSC, the number of microwave antennas at each iSC n (N_{mw}^n), the power for transmitting and receiving the aggregate backhaul traffic at each iSC (P_{link}^n) and the power consumption of switches at each iSC (P_{switch}^n) aggregating traffic from other iSCs in case more than one backhaul link originates at the reference iSC, as:

$$P_{Bh} = \sum_{n=1}^{N_{iSC}} P_{switch}^n (y_n) + N_{mw}^n P_{link}^n (y_n) \quad (10)$$

Note that the power consumption at any switch will depend on the aggregated traffic at the associated iSC and its

maximum capacity (Y_{max} in Gbps):

$$P_{switch}^n = \begin{cases} 0, & N_{mw}^n = 1 \\ & \text{or } y_n = 0 \\ P_s \left[\frac{y_n Y_{max} / f_{cell-Bh}}{C_{switch}} \right], & \text{otherwise} \end{cases} \quad (11)$$

where P_s stands for the switch basic power consumption and $f_{cell-Bh}$ is a factor denoting the relationship between backhaul traffic load and cell load.

Moreover, the power consumption for transmitting and receiving the aggregate backhaul traffic will generally depend on the traffic conditions. In this work, we consider that backhaul links can have an idle mode (i.e. when no data needs to be transported through backhaul) and a two-step function (low/high capacity traffic), where the two capacity regions are distinguished by a single threshold (C_{thr}):

$$P_{link}^n = \begin{cases} P_{idle}, & y_n = 0 \\ P_{low-traffic}, & 0 < y_n \leq \frac{C_{thr}}{Y_{max} f_{cell-Bh}} \\ P_{high-traffic}, & y_n \geq \frac{C_{thr}}{Y_{max} f_{cell-Bh}} \end{cases} \quad (12)$$

Table 3 presents the backhaul relevant parameters with exemplary realistic values. Note that backhaul traffic load is translated into cell load considering current LTE-based RAN. For this, we consider the iSC maximum bits-per-second capacity (evaluated assuming a single carrier with 10MHz bandwidth, 2×2 MIMO, 64QAM, and 28% control overhead) and the non-negligible overheads from X2 U- and C-plane (~4%), the transport protocol (~10%), and the IPsec (~14%) [21]. Finally, Figure 6 depicts the estimate backhaul power consumption where we can see that for generic small cells, the backhaul always operates in low capacity region, which results in flat power consumption for medium/high cell RF output power.

V. RANaaS-BASED OPTIMAL CONTROLLER FOR ENERGY EFFICIENT SMALL CELLS

In this section, the proposed power model is used to evaluate a mechanism that manages the activity of N_{iSC} iSCs according to access characteristics and the Quality of Service (QoS). We consider that part of the small cell functionalities, i.e.

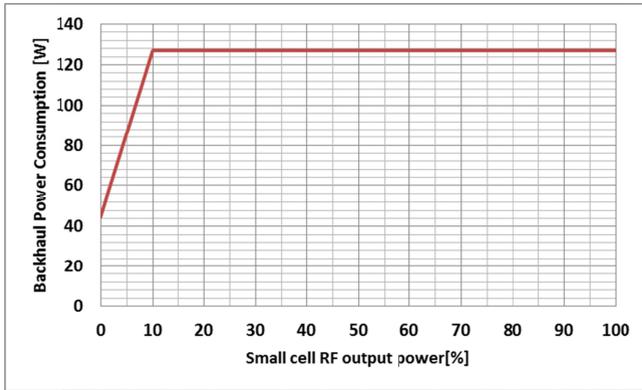


FIGURE 6. Backhauling Network Power consumption with respect to the small cell RF output power.

the Radio Resource Control, is shifted to the RANaaS. Accordingly, the RANaaS manages the activity of the iSCs to match available resources and required services.

Markov decision process is used to model this optimization problem [22]. In the considered scenario, the RANaaS receives data from the core network and stores it in a dedicated buffer. When required, RANaaS activates a given iSC through the backhaul and forwards the associated data to it. Thereafter, the iSC will autonomously manage available resources to transmit received data according to a first-in-first-out policy. When a given iSC is idle, energy saving is realised due to the reduced running functionalities at the small cell and at the backhaul network. On the contrary, as described in Section IV, when an iSC is activated, its contribution to the energy consumption of the RAN (that includes the backhaul) depends on its load, i.e., the transmitted data on the backhaul and access resources. Finally, when an iSC is deactivated, basic coverage and synchronization is provided to the nearby users by the macro cell through *dual connectivity* [23]. Accordingly, the iSCs can rely on *new carrier type*, which avoids systematic activation to broadcast cell-specific reference signals [23]. In this case, deploying the RANaaS at the macro cell site may enable to manage the status of a limited number of neighboring iSCs, which seems to be the optimal choice in terms of complexity and overhead.

In our model, the RANaaS is equipped with N_{iSC} buffers of size M packets, each one dedicated to an iSC. The number of packets received at an iSC at time step $t(f_t)$ is finite and depends on its load. Moreover, when activated, an iSC i transmits at most R_i packets, according to the available bandwidth and the transmission spectral efficiency.

Let S be a set describing the network state space and defined as $S = Q \times R \times F$, where Q , R , and F are the composite spaces modelling the buffer, the data rate and the incoming traffic, respectively. At each time step, the RANaaS observes the current state of the network $s_t \in S$ and selects an action from the action space A , where each action corresponds to activating at most one iSC and keep the others idle. The

decided action changes the current status from s_t to s_{t+1} , with a state transition probability $T(s_{t+1}|s_t; a_t)$, and incurs in a cost $C_t = C(s_t, a_t)$. Our goal is to find an optimal policy π that associates an action $a_t(s_t|\pi)$ to the state s_t . This policy aims to minimize the energy consumption and satisfying the QoS constraints while avoiding concurrent access of multiple iSCs. This problem can be represented as follows:

- a set of states S ;
- a set of actions A ;
- a state transition probability T ;
- a cost function C .

The statistics of the data rate and the incoming traffic are modelled as Markov chains, independent of the time steps, buffer state and actions. Hence, the model is Markovian because the state transition depends only on the current status/actions and is independent of any previous states/actions. Furthermore, we can define the system cost function as:

$$C_t = \tilde{P}(s_t, a_t) + \alpha \cdot \tilde{d}(s_t, a_t), \quad (13)$$

where α is a weighting factor that prioritizes between EE and QoS while $\tilde{P}(s_t, a_t)$ and $\tilde{d}(s_t, a_t)$ denote the sum of the N_{iSC} length vectors indicating the power consumption at iSCs and packet losses, respectively. Eq. (13) models the trade-off between the packet loss and the power consumption; to limit packet drop the RANaaS should keep the iSCs activated and transmit packets stored in its buffer. On the contrary, minimizing power consumption requires to keep iSCs idle, which in turn may lead to packet loss.

To provide performance evaluation, we consider a hotspot composed by four iSCs ($N_{iSC} = 4$) coordinated by the nearby RANaaS. Without loss of generality, we consider that in a transmission time interval (set equal to 1ms), at most one packet is received at each buffer and that an activated iSC can transmit up to two packets. Packet length is considered equal to 1Kbits. The model presented in Section IV is used to compute the network power consumption, due to the RANaaS platform, the backhaul and the iSCs. Accordingly, we assume that iSC and backhaul equipment in idle mode consumes the 60% of its zero-load power. Moreover, we consider that 5% of the BB load is transferred by the iSCs to the RANaaS to manage their activation/deactivation.

Figure 7 shows the performance of the optimal stationary solution (obtained through value iteration [22]) with respect to a random policy and greedy approach. By using the random policy, at each time step, the RANaaS randomly decides whether to keep iSCs idle or activated; on the contrary, when using the greedy method, action is taken to minimize the instantaneous cost, i.e., without considering the total (over time) expected cost. Solid, dotted and dashed lines, respectively, correspond to the optimal, greedy and random solutions. Performance is presented in terms of cumulative network EE [bit/J] computed as the ratio of the cumulative number of transmitted bits and the associated network power consumption measured over 1 second.

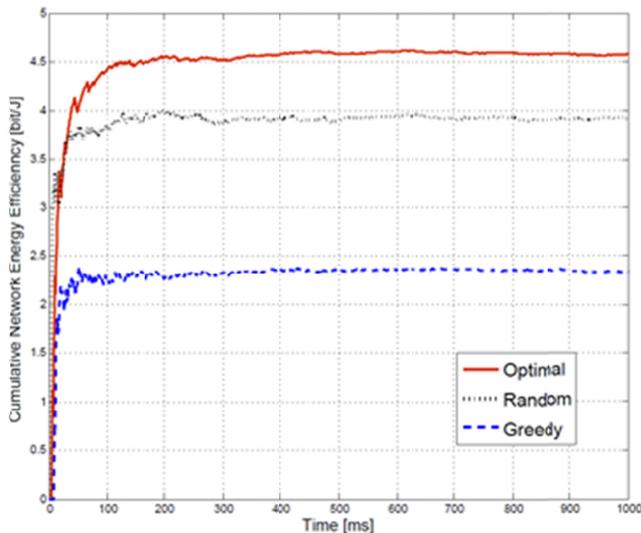


FIGURE 7. Cumulative Network EE with respect to different small cell management schemes.

The optimal solution leads up to 96% of gain with respect to the greedy solution; however, it gains only 16.4% with respect to the random policy. This surprising result is due to the limited energy consumption associated with the random solution, which in turns results in an unacceptable number of dropped data.

VI. CONCLUSIONS

This article discusses EE aspects and related benefits of RANaaS implementation (using commodity hardware) as evolution of LTE-Advanced networks toward 5G infrastructure. In fact, in the view of future definition of 5G systems, this cloud-based design is an important architectural solution, especially in terms of efficient usage of network resources. After providing a vision of the advantages of RANaaS and its benefits in terms of EE, we presented a consistent power model as a useful reference for further evaluations of flexible centralization of RAN functionalities in terms of EE. The incremental benefits through the years have been also discussed in perspective, by considering technological evolution of IT platforms and the increasing matching between their capabilities and the need for progressive virtualization of RAN functionalities. The description is finally complemented by an exemplary evaluation, in terms of EE, of a small cell activity controller based on the RANaaS paradigm, as a promising technology for future sustainable 5G radio.

ACKNOWLEDGMENT

The European Union and its agencies are not liable or otherwise responsible for the contents of this document; its content reflects the view of its authors only. The authors gratefully recognize the great contributions of many colleagues from iJOIN, who in fruitful cooperation, contributed with valuable insight, surveys and vision.

REFERENCES

- [1] *Code of Conduct on Energy Consumption of Broadband Equipment, Version 5.0*, European Commission, Brussels, Belgium, Dec. 2013.
- [2] A. D. Domenico, E. C. Strinati, and A. Capone, "Enabling green cellular networks: A survey and outlook," *Comput. Commun.*, vol. 37, pp. 5–24, Jan. 2014.
- [3] 3GPP, "Potential solutions for energy saving for E-UTRAN (release 11)," Radio Access Netw. Tech. Specification Group, Tech. Rep. 36.927, Sep. 2012. [Online]. Available: <http://www.3gpp.org>
- [4] 3GPP, "Solutions for energy saving within UTRA node B (release 11)," Radio Access Netw. Tech. Specification Group, Tech. Rep. 25.927, Sep. 2012. [Online]. Available: <http://www.3gpp.org>
- [5] 3GPP, "Telecommunication management; energy saving management (ESM); concepts and requirements (release 11)," Tech. Specification Group Services Syst. Aspects, Tech. Rep. TS 32.551, Jun. 2013. [Online]. Available: <http://www.3gpp.org>
- [6] *Green Meter Research Study: Reducing the Net Energy Consumption in Communications Networks by up to 90% by 2020*, GreenTouch, Kolkata, India, Jun. 2013.
- [7] I. Chih-Lin, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: A 5G perspective," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 66–73, Feb. 2014.
- [8] K. Sundaresan, M. Y. Arslan, S. Singh, S. Rangarajan, and S. V. Krishnamurthy, "FluidNet: A flexible cloud-based radio access network for small cells," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, Sep. 2013, pp. 99–110.
- [9] P. Rost et al., "Cloud technologies for flexible 5G radio access networks," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 68–76, May 2014.
- [10] "Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description; (release 11)," Radio Access Netw. Tech. Specification Group, Tech. Rep. TS 36.300, Sep. 2014. [Online]. Available: <http://www.3gpp.org>
- [11] National Industry of Standards and Technologies. *Cloud Computing Synopsis and Recommendations*. [Online]. Available: <http://csrc.nist.gov/publications/nistpubs/800-146/sp800-146.pdf>, accessed May 2012.
- [12] *HP Moonshot System*. [Online]. Available: <http://h17007.www1.hp.com/us/en/enterprise/servers/products/moonshot>, accessed Apr. 2014.
- [13] EARTH Project. *D2.3—Energy Efficiency Analysis of the Reference Systems, Areas of Improvements and Target Breakdown*. [Online]. Available: <http://www.ict-earth.eu/publications/publications.html>
- [14] G. Auer et al., "How much energy is needed to run a wireless network?" *IEEE Commun. Mag.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.
- [15] H. Holtkamp, G. Auer, V. Giannini, and H. Haas, "A parameterized base station power model," *IEEE Commun. Lett.*, vol. 17, no. 11, pp. 2033–2035, Nov. 2013.
- [16] C. Desset et al., "Flexible power modeling of LTE base stations," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2012, pp. 2858–2862.
- [17] CT FP7 FIT4GREEN. (Mar. 2012). *Presentation of the Full-Featured Federated Energy-Consumption Models*. [Online]. Available: http://www.fit4green.eu/sites/default/files/attachments/documents/D3.3_final.pdf
- [18] T. Werthmann, H. Grob-Lipski, and P. Proebster, "Multiplexing gains achieved in pools of baseband computation units in 4G cellular networks," in *Proc. IEEE 24th Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 3328–3333.
- [19] *Intel Xeon Processor 5500 Series*. [Online]. Available: http://download.intel.com/support/processors/xeon/sb/xeon_5500.pdf, accessed Apr. 2014.
- [20] P. Monti, S. Tombaz, L. Wosinska, and J. Zander, "Mobile backhaul in heterogeneous network deployments: Technology options and power consumption," in *Proc. 14th Int. Conf. Transparent Opt. Netw. (ICTON)*, Jul. 2012, pp. 1–7.
- [21] NGMN Alliance. *Small Cell Backhaul Requirements*. [Online]. Available: http://www.ngmn.org/uploads/media/NGMN_Whitepaper_Small_Cell_Backhaul_Requirements.pdf, accessed Jun. 2012.
- [22] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [23] C. Hoymann, D. Larsson, H. Koorapaty, and J.-F. Cheng, "A lean carrier for LTE," *IEEE Commun. Mag.*, vol. 51, no. 2, pp. 74–80, Feb. 2013.



DARIO SABELLA received his Dr. Eng. Degree in Electronics Engineering from Politecnico di Torino, Italy, in 2002. After graduation, he joined the research centre of Telecom Italia, called TILab, being involved in several R&D projects targeting future generation communications. In 2004 he received an additional post-degree specialization Master in Telecommunications. Currently working in the Wireless Innovation Department of Telecom Italia as responsible for some international projects

and innovative activities on Energy Efficiency for the mobile network. He is leading Telecom Italia activities for 5GrEEEn project, on green guidelines for 5G mobile system. From 2012 he is involved in Carbon Footprint project with the Italian Environment Ministry, and he is leading Telecom Italia activities related to energy efficiency of the mobile network in the framework of the LTE auction in Italy. He is co-author of several publications and patents in the field of wireless communications and energy efficiency.



ANTONIO DE DOMENICO received the M.Sc. degree in telecommunication engineering from the University of Rome La Sapienza, Rome, Italy, in 2008, and the Ph.D. degree from the University of Grenoble, Grenoble, France, in 2012. Since 2009, he has been with the Laboratory for Electronics and Information Technology, Micro and Nanotechnology Innovation Centre, Atomic Energy Commission, Grenoble, as a Research Engineer. His current research topics are cloud-

enabled heterogeneous wireless networks, millimetric-wave-based communications, machine learning, and green communications. He is the main Inventor or Co-Inventor of seven patents.



EFSTATHIOS KATRANARAS (M'06) received the B.Sc. degree in electrical and computer engineering from the National Technical University of Athens, Athens, Greece, in 2005, and the M.Sc. (Hons.) and Ph.D. degrees in mobile communications from University of Surrey, Guildford, U.K., in 2006 and 2009, respectively. He is currently a Research Fellow with the Institute for Communication Systems, University of Surrey, where he is an Active Research Member of the

5G Innovation Centre. His main research interests include the evaluation of fundamental capacity limits and energy efficiency in cellular systems and advanced multicell cooperation techniques for future dense small cell networks.



MUHAMMAD ALI IMRAN (M'03–SM'12) received his M.Sc. (Distinction) and Ph.D. degrees from Imperial College London, UK, in 2002 and 2007, respectively. He is currently a Reader in the Institute for Communication Systems, home of 5G Innovation Centre, at the University of Surrey, UK. In this role, he is leading a number of multimillion international research projects encompassing the areas of energy efficiency, fundamental performance limits, sensor networks and self-organising

cellular networks. He is also leading the new physical layer work area for 5G innovation centre at Surrey. He has a global collaborative research network spanning both academia and key industrial players in the field of wireless communications. He has supervised 18 successful PhD graduates and published over 150 peer-reviewed research papers including more than 20 IEEE Journals. His research interests include the derivation of information theoretic performance limits, energy efficient design of cellular system and learning/self-organising techniques for optimisation of cellular system operation. He is an editor of *IET Communications* and a guest editor of special issues in *IEEE COMMUNICATIONS MAGAZINE* and *IEEE WIRELESS COMMUNICATIONS MAGAZINE*. He is a senior member of IEEE and a Fellow of Higher Education Academy (FHEA), UK. He has won the Faculty Learning and Teaching Award 2014 and has been awarded the 2014 IEEE Communications Society Fred W. Ellersick Prize.



MARCO DI GIROLAMO Graduated in Electronics Engineering (*summa cum laude*) in November 1986. More than 25 years of working experience in R&D, IT and Consulting environments, in different Companies, both Italian and international. In HP since 1993, covered positions in R&D as Development Engineer, in Customer Care and IT Department. Since 2000, he is working in the Italy Innovation Centre, where served as senior technology specialist, solution architect and project

manager in different engagements, both internal and external.

Since 2002, he is leading HP Italy activities into European collaborative research and innovation. In this position, he coordinated and led HP Italy initiatives related to European innovation programs, including IST FP5, FP6, FP7, eContent and CIP project participations. He co-authored different publications within the projects he contributed to. In particular, he contributed as WP leader to the FIT4Green and All4Green projects, which investigated methods, tools and metrics to improve energy efficiency of datacentres in integrated ecosystems. He is currently engaged in the DC4Cities project, investigating technologies to maximize the share of renewable energy used in supplying datacentres.



UMER SALIM received the Ph.D. and M.S. degrees in electrical engineering with a specialization in communication theory and signal processing from EURECOM, Biot, France, and Supélec, Cesson-Si-vignif, France. He has several years of research experience in digital communications and signal processing, and has authored several papers in well-known conferences and journals. His main areas of interest include signal processing techniques for multicell multiuser multiple-input

multiple-output systems, novel and practical CSI feedback design techniques and analysis, information theoretic analysis of cognitive radio, and multiuser information theory in general. He has been serving as a reviewer of the *IEEE TRANSACTIONS ON INFORMATION THEORY*, the *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, the *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, and numerous well-known conferences. He co-authored a paper that received the Best Paper Award at the European Wireless Conference in 2011.

He is currently with the Department of Systems Engineering, Intel Mobile Communications, Paris, France, where he is involved in the design of advanced receivers for future wireless standards and has designed sophisticated interference cancellation algorithms, which are in use in the modern high-end smart-phones and tablets.



MASSINISSA LALAM received the master's degree from the Institut National Polytechnique de Grenoble, école Publique D'ingénieurs en Informatique, Mathématiques Appliquées et Télécommunications de Grenoble, Grenoble, France, in 2002, and the Ph.D. degree from the Ecole Nationale Supérieure de Télécommunications de Bretagne, Telecom Bretagne, Plouzané, France, in 2006, where he held a post-doctoral position, in 2007. He joined

Orange Labs, Paris, France, from 2008 to 2009. He is currently an Expert with Sagemcom Broadband, Paris, France, where he works on wireless (Wi-Fi) and cellular (3G/LTE) technologies. He has been involved in several European projects, such as BeFEMTO, iJOIN, and LEXNET, as a Work Package Leader and Task Leader or Contributor. His expertise includes in link and system-level performance evaluation, heterogeneous network, network modeling, and radio resource management. He has authored or co-authored over 15 international publications.



KONSTANTINOS SAMDANIS received the Ph.D. and M.Sc. degrees in mobile communications from Kings College London, London, U.K. He is currently a Senior Researcher and Backhaul Standardization Specialist with NEC Laboratories Europe, Heidelberg, Germany. He leads a research project on long-term evolution (LTE) network management and multitenancy, while at the same time, he is involved in the Marie Curie Initial Training Network through the CROSSFIRE project as a

Work Package Leader on LTE network virtualization. He is currently active in 5G, while in the past, he participated in the iJOIN and BeFemto FP7 projects and IST EVEREST project. He is also active in the Broadband Forum in Software-Defined Networking/Network Functions Virtualization and is an Editor of the *TR-293 Energy-Efficient Mobile Backhaul*. He has served as a Guest Editor of the IEEE COMMUNICATION MAGAZINE and the IEEE MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE E-LETTERS, and the Co-Chair of the Next-Generation Networking Symposium of the IEEE International Conference on Communications 2014.



ANDREAS MAEDER received the Ph.D. degree from the University of Würzburg, Würzburg, Germany, in 2008, with his thesis on radio resource management in integrated 3GPP UMTS/HSPA systems. Since 2008, he has been with NEC Laboratories Europe, Heidelberg, Germany, where he is currently a Senior Researcher working on 5G mobile networks, including RAN and core network virtualization, energy efficiency, and the convergence of IT and telecommunication technologies. He has contributed to the standardization of mobile networks in IEEE and 3GPP since 2008. He has authored numerous scientific articles and conference papers, holds patents, and is an Editor of a book on green data communication.

• • •